Segmentation of brain lesions from CT images based on deep learning techniques

Xiaohong Gao, Yu Qian*
Department of Computer Science, Middlesex University, UK. x.gao@mdx.ac.uk
*Cortexica Vision Systems, London, UK. yuqian_hong@yahoo.com

Abstract: While Computerised Tomography (CT) may have been the first clinical tool to study human brains when any suspected abnormality related to the brain occurs, the volumes of CT lesions usually are usually disregarded due to variations among inter-subject measurements. This research responds to this challenge by applying the state of the art deep learning techniques to automatically delineate the boundaries of abnormal features, including tumour, associated edema, head injury, leading to benefiting both patients and clinicians in making timely accurate clinical decisions. The challenge with the application of deep leaning based techniques in medical domain remains that it requires datasets in great abundance, whilst medical data tend to be in small numbers. This work, built on the large field of view of DeepLab convolutional neural network for semantic segmentation, highlights the approaches of both semantics-based and patch-based segmentation to differentiate tumour, lesion and background of the brain. In addition, fusions with a number of other methods to fine tune regional borders are also explored, including conditional random fields (CRF) and multiple scales (MS). With regard to pixel level accuracy, the averaged accuracy rates for segmentation of tumour, lesion and background amount to 82.9%, 85.7%, 85.3% and 81.3% while applying the approaches of DeepLab, DeepLab with MS, DeepLab with MS and CRF, and patch-based pixel-wise classification respectively. In terms of the measurement of intersection over union of two regions, the accuracy rates are of 70.3%, 75.1%, 77.2%, and 63.6% respectively, implying overall DeepLab fused with MS and CRF performs the best.

Key words: deep learning, segmentation, classification, CT brain tumours, brain lesions, DeepLab.

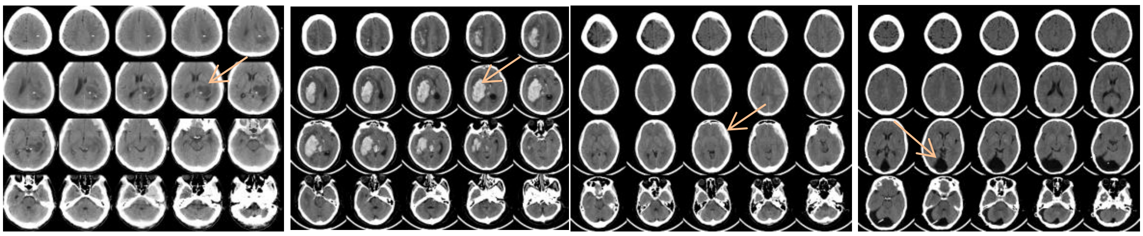
1. Introduction

Computerised Tomography (CT) is the first imaging tool to study the brain and remains the first clinical scanner to undertake when any suspected abnormality in the brain, e.g., a tumour, occurs due to its prevalent, economical and easy to operate nature. The outcome of CT scans will then be applied to determine subsequent treatment planning. In the case of a tumour, not only its type and location, which can be ascertained by the digital brain Atlas and the procedure of biopsy but also its volume play a crucial role in making this clinical decision, for instance, whether to perform chemotherapy, or undertake Magnetic Resonance (MR) scan for further confirmation, or proceed with neurosurgery. Furthermore, accurate measurement of tumour size will also assist to establish the effectiveness of the treatment by assessing whether tumour decreasing or spreading under a certain treatment. While a CT image depicts clear structural information of the brain, it does not show boundaries as clearly as an MR image due to its low resolution at tissue level. In this work, the state of the art deep learning techniques are exploited to segment brain tumours and lesions (head injury, bleeding and swelling), in an attempt to realise accurate measurement automatically, leading to making the most of this valuable first-hand CT data to benefit both patients to receive timely treatment and clinicians in shortening prolonged tests.

Segmentation of CT brain images has been conducted by a number of researchers applying clustering approach [1]. However, those work mainly has a focus on the structure of the brain, such as cerebrospinal fluid (CSF) and brain matter. In particular, the delineation of the boundary of abnormal regions is only mentioned and less fully addressed. On the other hand, accurate measurement can contribute significantly to the diagnosis of brain diseases, for example, Alzheimer's [2]. Therefore, CT measurements can be of great value to the work-up of decision making processes for the brain.

This work will fill this gap by automatically segmenting lesion sizes/volumes by employing the state of the art of convolutional neural network (CNN). The challenge with processing medical images with CNN remains that medical images tend to have small datasets while CNN works better with more training data, for instance, in millions in the training of DeepLab [3], which will be addressed in the this work.

Figure 1 illustrates four samples of various types of lesions (arrows on representative slices) in axial direction in this collection.



Glioma Tumour Bleeding Head Injury

Figure 1. Examples of lesioned brain CT images pointed by arrows on selected slices of 3D datasets.

2. Methodology

2.1 Datasets

Similar to [4], 115 lesioned datasets of 3D are collected from Navy General Hospital in China. Upon application, they are processed in 2D form, i.e. slice by slice rather than taking a dataset as a whole, since each dataset has relatively low resolution in z-direction (~0.5mm depth whereas a typical MR has ~0.1-0.2mm) and only contain a few lesioned slices (Figure 1). Another reason of applying 2D form instead of 3D is to increase dataset sizes. In total, 355 slices are selected. For each lesioned slice, a mask is created by a clinician to delineate lesion boundaries. For the tumours, in addition to their borders, the boundaries of edema are also delineated if there are any.

2.2 Convolutional neural network (CNN) and DeepLab

Deep learning models refer to a class of computing machines that can learn a hierarchy of features by building high-level attributes from low-level ones [5, 6] , thereby automating the process of feature construction. One of these models is the well-known convolutional neural network (CNN) [7]. Consisted of a set of algorithms in machine learning, CNN, stemming from human vision theory, comprises several (deep) layers of processing involving learnable operators (both linear and non-linear), and hence has the ability to learn and build high-level information from low-level features in an automatic fashion [8], hence achieving start of the art results in a number of computer vision fields, in particular in classification, localisation and segmentation.

Since CNN based techniques apply the convolutional and pooling layers (down-sampling) to represent high-level features of input images, for segmentation, while the same CNN networks used for classification are applied, extra steps are in need to achieve pixel-level accuracy, including converting fully connected layer to convolutional layer and up-sampling using a deconvolutional layer. Another direction of research for pixel-wise prediction focuses on patch-based classification. Each pixel of input image is labelled by classifying the cropped patch centred on this pixel. In this way, a classifier is trained by using CNN on a set of cropped patches which locate on different classes of mask, edges of masks and background as well.

To achieve accurate results, fused CNN networks are usually in need with the inclusion of refinement methods. Currently, the state of the art refinement approaches include superpixel [9] and Conditional Random Field (CRF) [10]. Superpixel intends to over segment, aiming at generating a coherent grouping of pixels, provides a convenient way to compute image features and reduces the complexity of subsequent image processing tasks [11]. Whereas CRF, through the application of a class of statistical modelling methods tends to provide structured prediction, hence achieving refinement of segmentation results.

For instance, DeepLab [3, 12], a fully convolutional network (FCN) with integration of CRF, has achieved the best segmentation results on the challenge of PASCAL VOC 2012 images. The advantage of DeepLab over the other CNN networks is that DeepLab introduces an atrous (with holes) algorithm [13] to speed up the segmentation of feature maps by replacing the deconvolutional layers through the introduction of zeros to convolutional filters to increase their length (up-sampling) after the last two max-pooling layers. In this way, the Field-of-View (FOV) of the models can be arbitrarily controlled by adjusting the input stride, without increasing the number of parameters or the amount of computation.

In this study, DeepLab framework [3] is investigated on segmentation of CT brain images through the fusion with Fully Convolutional Networks (FCN) [14] to realise pixel-wise prediction and Conditional Random Field (CRF) for definition of edges. In doing so, the pre-trained dCNN classifier [15] is firstly converted by replacing the last fully connected layers with fully convolutional layers to produce coarse output maps. Then, upon those blob-like coarse segmentations, a fully-connected pairwise CRF is applied to fine tune edge details. This is done by assembling neighbouring nodes by assigning same labels to spatially proximal pixels, which is realised by the energy function as formulated by Eq. (1).

$$E(\boldsymbol{x}) = \sum_i \theta_i(x_i) + \sum_{ij} \theta_{ij}(x_i, x_j) \quad (1)$$

Where
$$\theta_i(x_i) = -\log P(x_i) \quad (2)$$

and

$$\theta_{ij}(x_i, x_j) = \mu(x_i, x_j)\left[w_1 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2}\right) + w_2 \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2}\right)\right] \quad (3)$$

In addition, $P(x_i)$ refers to the output from FCN, being the probabilistic label assignment at pixel i, and $\theta_{ij}(x_i, x_j)$ the pairwise potential defined in terms of colour vector I_i and I_j and positions of p_i and p_j.

In this research, two FCN models are trained on 309 CT images and evaluated on 46 testing images. All abnormal CT images were manually segmented to tumour and lesion by a medical doctor. To contend with the shortage of CT datasets, this study applies a pre-trained model of VGG-16 [15] on ImageNet. Specifically, the first FCN model with large field-of-view whereby sampling rate is set to 12 (i.e. $r = 12$), is trained for pixel labelling. Then the second FCN model is fine-tuned with multiple scales (MS) built upon the first model. Finally, CRF is annexed to the top of second FCN model to define edge details with parameter settings in Eq. (3) as: $w_1 = 5$; $w_2 = 1$; $\sigma_a = 10$; $\sigma_\beta = 1$; and $\sigma_\gamma = 3$.

3. Results

Figure 2 illustrates a number of examples of segmentation results applying the extended network of DeepLab fused with MS and CRF where red colour indicating the region being tumour or bleeding and green the lesions (e.g. edema). Quantitatively, the comparison results are presented in Table 1, including DeepLab, DeepLab with MS, DeepLab with MS and CRF) and patch based segmentation respectively, where DeepLab refers to with large FOV. With regard to patch-based segmentation approach, the simple AlexNet, a built-in model in Matlab software, is applied for patch classification. The training patches are rendered firstly by cropping regions to 64×64 pixel and then randomly selected with central pixels located in the masks of tumours and lesions, as well as edge of mask and background, similar to [16]. After merging edge patches to background, the training dataset with 3 labels (i.e. tumour, lesion and background) are generated. The classification accuracy for 3 classes is 81.3%. For pixel wise prediction, each pixel of test image is labelled by classifying the cropped patch centred on this pixel. For all four methods, the measurement of the evaluation is conducted by using Intersection over Union (IoU) and Pixel Accuracy (PA) on each segmented class (Tumour, Lesion and Background) and the mean value of each class as formulated in Eqs. (4) and (5).

$$IoU = \frac{\text{True positive}}{\text{True positive+False positive+False Negative}} \tag{4}$$

$$PA = \frac{\text{True positive}}{\text{True positive+False Negative}} \tag{5}$$

Table 1. The comparison result between varying fused CNN network where DeepLab indicates with large field of view, i.e. γ=12.

Method		**Tumour (%)**	**Lesion (%)**	**Background (%)**	**Mean PA (%)**	**Mean IoU (%)**
DeepLab	PA	85.10	64.75	98.88	82.91	
	IoU	61.18	52.03	97.73		70.31
DeepLab + MS	PA	88.51	69.69	99.03	**85.74**	
	IoU	70.01	57.31	97.97		75.10
DeepLab + MS + CRF	PA	88.07	68.65	99.24	85.32	
	IoU	75.02	58.44	98.12		**77.19**
Patch-based pixel-wise	PA	78.64	68.55	96.83	81.34	
	IoU	58.20	37.17	95.61		63.64

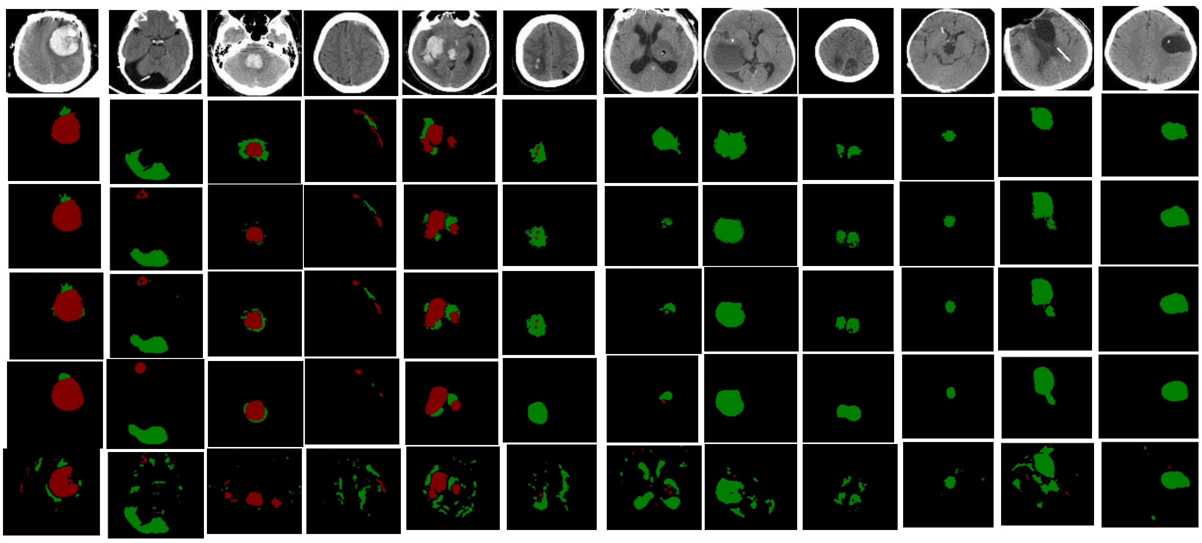


Figure 2. Tumour segmentation results with large FOV DeepLab and patch-based deep learning networks. From top to bottom: original image, ground truth, DeepLab +MS+CRF, DeepLab+MS, DeepLab and patch-based deep learning, where different colours indicate difference regional classes.

At pixel level, the best segmentation accuracy can be achieved by the approach of DeepLab coupled with multiple scales (MS), giving rise to the accuracy rate of 85.74%. All the methods perform well on segmentation of tumours with over 81% accuracy rate but less so for segmentation of lesions, which is expected as some lesion regions, e.g., bleeding in Figure 2 (row 4), are merging with health tissues. When judged based on IoU, the approach of DeepLab fused with MS and CRF delivers the best result, achieving 77.19% of accuracy rate.

While patch-based deep learning network performs better for segmentation of brain tumour on MR images [17], it appears to be less so for CT images. Partly due to the fact that these CT images have not undergone pre-processing stage to normalise intensity levels across all CT images as in the case of MR images. As illustrated in Figure 2, intensity distribution across different datasets varies considerably. On the other hand, application of multi-scale to DeepLab can alleviate this challenge by increasing the accuracy of boundary localization, which is realised through the attachment of two layers of multiple layer perceptrons (MLP) to the input image and the output of each of the first four max pooling layers.

4. Conclusion and discussion

This research evaluates the current state of the art deep learning (DL) techniques for segmentation of brain lesions from CT images. It showcases that when coupled with fine-tuned techniques, e.g. CRF and MS, deep learning based approach can provide accurate and robust results without too much involvement of pre-processing work. These findings will be taken forward in the future to the application to diagnosis of early onset of Alzheimer's disease (AD) applying deep learning based segmentation approaches to measure atrophy factors between temporal horn radio and suprasellar cistern ratio, leading to revealing significant insights of AD. While the first scan of CT images undertaken by patients may have shown AD signature features, the manual delineation of atrophy in certain regions, for example, medial temporal lobe and left hippocampus, vary considerably between radiologists [2].

In medical field, patched-based segmentation is also widely applied, which appears to perform better in a number of modalities, e.g. MR or microscopy. In this study, patch-based segmentation techniques applying DL, as described in [16], is also evaluated, which delivers the average PA of 81.34%, less than the averaged PA (84.66%) of all 4 approaches applied in this work. At present, the 3D CT images are treated in 2D form, slice by slice. In the future, more datasets with higher resolution in z-direction will be collected upon which 3D based DL approaches of segmentation will be developed to take advantage of the information along depth-direction. To overcome the shortage of medical datasets, pre-trained network, e.g. VGG-16 [15] and Alexnet built on ImageNet, has been applied to lay a foundation for training, which appears to work well even this model is built upon natural images, e.g. bicycle, human, boat, etc..

Acknowledgement

This work constitutes part of project WIDTH that is funded by EU FP7 under Marie Curie Scheme (IRSES, No 269124). Their financial support is gratefully acknowledged. The authors would also like to thank Dr. Alice Gao from Southend University Hospital NHS Foundation Trust at the UK for helping creating masks of brain lesions.

References:

[1] Lee T., Fauzi M., Komiya R. , Segmentation of CT brain images using unsupervised clusterings, Journal of Visualization, Volume 12 (2): 131–138, 2009.

[2] Zhang Y., Londos E., Minthon L., Wattmo C., Liu H., and Wahlund L.O. Usefulness of computerd tomotgraphy linear measurements in diagnosing alzheimers disease. Acta Radioligica, 49(1):91–97, 2008.

[3] Chen, L., Papandreou G., Kokkinos I., Murphy K., Yuille A., DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, TPAMI, *accepted* 2017. https://arxiv.org/abs/1606.00915. Retrieved in June 2017.

[4] Gao, X., Hui R., Tian Z., Classificaiton of CT images based on deep learning network, *Computer Methods and Programs in Biomedicine*, 138:49-56, 2017.

[5] Fukushima K., Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biol. Cyb.*, 36: 193–202, 1980.

[6] LeCun Y., Bottou L., Bengio Y., and Haffner P., Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86(11): 2278–2324, 1998.

[7] LeCun Y., Huang F.J. , Bottou L., Learning methods for generic object recognition with invariance to pose and lighting, *Processings of Computer Vision and Pattern Recognition (CVPR)*, 2: II-97-104, 2004.

[8] LeCun Y., Bengio Y., Hinton G., Deep Learning, *Nature, 521: 436-444, 2015.*

[9] A. Lucchi, Y. Li, X. Boix, K. Smith, and P. Fua, "Are spatial and global constraints really necessary for segmentation?" in ICCV, 2011.

[10] Krähenbühl P., and V. Koltun, Efficient inference in fully connected CRFs with gaussian edge potentials, in NIPS, 2011.

[11] Shen J., Du Y., Wang W., Li X.. *Lazy random walks for superpixel segmentation*. Transactions on Image Processing 23 (4) (2014) 1451-1462.

[12] Jia Y., Shelhamer E. Donahue J, Karayev S. and Long, J. and Girshick, Caffe: Convolutional Architecture for Fast Feature Embedding, arXiv, 2014.

[13] Rother C., Kolmogorov V., and Blake A., GrabCut: Interactive foreground extraction using iterated graph cuts, SIGGRAPH, 2004.

[14] Long J., Shelhamer E., and T Darrell., Fully convolutional networks for semantic segmentation, CVPR, 2015.

[15] Simonyan K. and Zisserman A., Very deep convolutional networks for large-scale image recognition, ICLR, 2015.

[16] Janowczyk A. and Madabhushi A., Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases, *Journal of Pathology Informatics*, 7(1):1-29, 2016.

[17] Pereira S., Pinto A., Alves V., and Silva C., Brain tumour segmentation using convolutional neural networks in MRI images, *IEEE transactions on medical imaging*, 35(5):1240-1251, 2016.